

CLAIMS

1. A method of controlling high-speed reading in a text-to-speech conversion system including a text analysis module for generating a phoneme and prosody character string from an input text; a prosody generation module for generating a synthesis parameter of at least a voice segment, a phoneme duration, and a fundamental frequency for said phoneme and prosody character string; a voice segment dictionary in which voice segments as a source of voice are registered; and a speech generation module for generating a synthetic waveform by waveform superimposition by referring to said voice segment dictionary,

said method comprising the step of providing said prosody generation module with a phoneme duration determination unit that includes both a duration rule table containing empirically found phoneme durations and a duration prediction table containing phoneme durations predicted by statistical analysis and determines a phoneme duration by using, when a user-designated utterance speed exceeds a threshold, said duration rule table and, when said threshold is not exceeded, said duration prediction table.

2. The method according to claim 1, wherein said threshold is a predetermined maximum utterance speed.

3. A method of controlling high-speed reading in a text-to-speech conversion system including a text analysis module for generating a phoneme and prosody character string from an input text; a prosody generation module for generating a synthesis parameter of at least a voice segment, a phoneme duration, and a fundamental frequency for the phoneme and prosody character string; a

voice segment dictionary in which voice segments as a
source of voice are registered; and a speech generation
module for generating a synthetic waveform by waveform
superimposition while referring to said voice segment
5 dictionary,

said method comprising the step of providing said
prosody generation module with a pitch contour
determination unit that has both an empirically found rule
table and a prediction table predicted by statistical
10 analysis and determines a pitch contour by determining both
accent and phrase components with, when a user-designated
utterance speed exceeds a threshold, said duration rule
table and, when said threshold is not exceeded, said
duration prediction table.

15 4. The method according to claim 3, wherein
said threshold is a predetermined maximum utterance speed.

5. A method of controlling high-speed reading in
a text-to-speech conversion system including a text
analysis module for generating a phoneme and prosody
20 character string from an input text; a prosody generation
module for generating a synthesis parameter of at least a
voice segment, a phoneme duration, and a fundamental
frequency for the phoneme and prosody character string; a
voice segment dictionary in which voice segments as a
25 source of voice are registered; and a speech generation
module for generating a synthetic waveform by waveform
superimposition by referring to said voice segment
dictionary,

said method comprising the step of providing said
30 prosody generation module with a sound quality coefficient
determination unit that has a sound quality conversion
coefficient table for changing said voice segment to switch
sound quality and selects from said sound quality

20251044018001

conversion coefficient table such a coefficient that sound quality does not change when a user-designated utterance speed exceeds a threshold.

5 6. The method according to claim 5, wherein said threshold is a predetermined maximum utterance speed.

10 7. A method of controlling high-speed reading in a text-to-speech conversion system including a text analysis module for generating a phoneme and prosody character string from an input text; a prosody generation module for generating a synthesis parameter of at least a voice segment, phoneme duration, and fundamental frequency for the phoneme and prosody character string; a voice segment dictionary in which voice segments as a source of voice are registered; and a speech generation module for
15 generating a synthetic waveform by waveform superimposition by referring to said voice segment dictionary,

20 said method comprising the step of providing said prosody generation module with both a pitch contour correction unit for outputting a pitch contour corrected according to an intonation level designated by the user and a switch for determining whether a base pitch is added to said pitch contour corrected according to said user-designated utterance speed.

25 8. The method according to claim 7, wherein said threshold is a predetermined maximum utterance speed.

30 9. The method according to claim 7, wherein said pitch contour correction unit performs a pitch contour generation process that includes a phrase component calculation process in which all phrases of an input sentence are processed by calculating a phrase component by statistical analysis according to said user-designated utterance speed or making said phrase component zero and a process in which all words in said input sentence are

10058104.012902

processed by calculating an accent component by statistical analysis according to said user-designated utterance speed and either correcting said accent component according to said user-designated intonation level or making said accent component zero.

10. A method of controlling high-speed reading in a text-to-speech conversion system including a text analysis module for generating a phoneme and prosody character string from an input text; a prosody generation module for generating a synthesis parameter of at least a voice segment, a phoneme duration, and a fundamental frequency for said phoneme and prosody character string; a voice segment dictionary in which voice segments as a source of voice are registered; and a speech generation module for generating a synthetic waveform by waveform superimposition while referring to said voice segment dictionary,

said method comprising the step of providing said speech generation module with signal sound generation means for inserting a signal sound between sentences to indicate an end of a sentence when a user-designated utterance speed exceeds a threshold.

11. The method according to claim 10, wherein said threshold is a predetermined maximum utterance speed.

12. A method of controlling high-speed reading in a text-to-speech conversion system including a text analysis module for generating a phoneme and prosody character string from an input text; a prosody generation module for generating a synthesis parameter of at least a voice segment, a phoneme duration, and a fundamental frequency for the phoneme and prosody character string; a voice segment dictionary in which voice segments as a source of voice are registered; and a speech generation

module for generating a synthetic waveform by waveform
superimposition by referring to said voice segment
dictionary,

5 said method comprising the step of providing said
prosody generation module with a phoneme duration
determination unit for performing a process in which when a
user-designated utterance speed exceeds a threshold, an
utterance speed of at least a leading word in a sentence is
returned to a normal utterance speed.

10 13. The method according to claim 12, wherein
said threshold is a predetermined maximum utterance speed.

15 14. The method according to claim 12, wherein
said phoneme duration determination unit performs a process
in which when a word under process is a leading word in a
sentence and said user-designated utterance speed exceeds
said threshold, a phoneme duration is not corrected and,
when said word under process is not a leading word of a
sentence or said user-designated utterance speed does not
20 exceed said threshold, a first process by which a phoneme
duration correction coefficient is changed according to
said user-designated utterance speed and a second process
in which all syllables of said word are processed by
correcting a length of a vowel or vowels of said word, and
carrying out said first and second processes for all words
25 contained in the sentence.

20058104.012900